The biomarker paradigm: between diagnostic efficiency and clinical efficacy

Giuseppe Lippi¹, Camilla Mattiuzzi²

¹ Laboratory of Clinical Chemistry and Hematology, Academic Hospital of Parma, Parma, Italy
² Service of Clinical Governance, General Hospital of Trento, Trento, Italy

ABSTRACT

The interest in biomarker research has been growing exponentially, and this trend is not expected to reverse soon. Although the clinical usefulness of laboratory tests is conventionally defined in terms of diagnostic efficiency or clinical efficacy (or effectiveness), these notions are complementary but not interchangeable. The former concept is an expression of diagnostic accuracy but does not entail outcome assessment. Conversely, clinical efficacy investigates whether or not a certain test can produce significant changes in managed care and an improvement of clinical outcomes. The vast majority of published studies were mainly focused on diagnostic efficacy rather than on clinical efficacy, and this seems no longer sustainable in a world with limited resources. Although bridging the gap between efficiency and efficacy is not a trivial endeavour, a paradigm shift is necessary, wherein the laboratory community should focus on what clinicians need rather than pursuing an endless search for analytically perfect tests. In the foreseeable future, efficacy should be improved by translating this concept into a simple “six R” paradigm, namely, performing the Right test, with the Right method, at the Right time, to the Right patient, at the Right cost, for the Right outcome.

Introduction  Laboratory diagnostics is conventionally defined as the process of identifying the nature and cause of human disorders by using in vitro testing. It is now generally acknowledged that up to 70% of clinical decisions are variably influenced by the results of diagnostic testing, so that this branch of science and medicine has an undisputed value in patient care. In fact, an improved diagnostics could lead to earlier detection of pathologies and appropriate intervention, thus improving health outcomes and minimizing the costs of disease and its complications.

The evidence of a strong association between biomarkers and disease resulted in an impressive number of studies in this field. If one considers the number of documents published per year in Scopus (ie, one of the most widely accessed scientific database) using the keyword “biomarkers” in article title, abstract, and keywords, the electronic search from the years 1977 to 2014 is described by an exponential curve, which is seemingly not flating even in the year 2013 (Figure 1). Interestingly, a total number of 83 013 documents could be retrieved, with an impressive 17-fold increase from 2000 to 2013 (ie, from 818 to 14 099), with a mean increase of 25% per year in the past decade. There are numerous factors that contributed to this impressive trend of growing scientific and clinical interest, including a major understating of disease mechanisms and individual patient responses to therapy, along with remarkable technological advances in laboratory techniques, which have progressed much faster than other diagnostic disciplines, such as diagnostic imaging, and simplified the measurement of a vast array of biomarkers.

Definitions  The Working Group of the National Institute of Health (NIH) has provided a reliable definition of a biomarker, as “a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”. According to this widely accepted definition, biomarkers can be potentially used throughout the natural history of a disease, and hence for risk assessment (eg, screening), diagnosis, prognostication, or therapeutic monitoring (Figure 2). Rather understandably, some biomarkers may be used at multiple steps, whereas
The biomarker paradigm: between diagnostic efficiency and clinical efficacy...

In the Encyclopedia Britannica, \(^3\) efficiency is defined, in literal terms, as “the ability to do or produce something without wasting materials, time, or energy”. Conversely, efficacy is defined as “the ability to produce a desired or intended result”. Effectiveness, which is seldom used as a surrogate of efficacy, is a direct measure of the former aspect, and is hence defined as “the degree to which something is successful in producing a desired or intended result”. It is hence obvious that while efficiency only expresses the quality (or degree) of being efficient, efficacy (or effectiveness) refers to a process or activity that is capable of modifying (favorably) one or more outcomes. Translating these terms into clinical practice, diagnostic efficiency is defined as the ability to identify or rule out the presence of a given disease, whereas diagnostic efficacy (or effectiveness) does not necessarily expresses the appropriateness of a diagnostic test in statistical terms, but is mainly aimed at evaluating whether or not a certain test can produce a significant change in managed care and ultimately provide an improvement in clinical outcomes. \(^4\)

Interestingly, although clinical efficiency mainly depends on whether a health care intervention is worth its cost to individuals or society, clinical efficacy defines the extent to which a health care intervention achieves its intended effect in the usual clinical setting. \(^3,4\)

Therefore, the diagnostic efficiency of a given biomarker is an expression of the accuracy to correctly identify a given (pathological) condition.

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Drawbacks of translating diagnostic efficacy into clinical efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>feeble relationship between biomarker research and the leading causes of death</td>
</tr>
<tr>
<td></td>
<td>lost in transition</td>
</tr>
<tr>
<td></td>
<td>inadequateness of derivation and validation cohorts</td>
</tr>
<tr>
<td></td>
<td>low diagnostic accuracy</td>
</tr>
<tr>
<td></td>
<td>poor analytical performance</td>
</tr>
<tr>
<td></td>
<td>low analytical sensitivity</td>
</tr>
<tr>
<td></td>
<td>errors or fraud in published data</td>
</tr>
<tr>
<td></td>
<td>high biological variability</td>
</tr>
<tr>
<td></td>
<td>technological issues</td>
</tr>
<tr>
<td></td>
<td>cost-effectiveness</td>
</tr>
<tr>
<td></td>
<td>inappropriateness</td>
</tr>
<tr>
<td></td>
<td>overregulation by regularity agencies</td>
</tr>
</tbody>
</table>

others only serve for a single phase. The cardio-specific troponins measured with highly sensitive (HS) immunoassays are paradigmatic examples of “multitasking” biomarkers, since they can now be reliably used for identifying patients with stable coronary artery disease, diagnosing acute coronary syndrome, establishing the prognosis after irreversible myocardial injury and establishing the effectiveness of revascularization therapy (FIGURE 2). Good examples of biomarkers which only target one step throughout the natural history of cardiovascular disease include cholesterol (risk assessment), ischemia-modified albumin (diagnosis), natriuretic peptides (prognostication), and platelet function testing (monitoring the effectiveness of antiplatelet therapy).

FIGURE 1 Number of documents published per year in Scopus using the keyword “biomarkers” in article title, abstract, and keywords, from the years 1977 to 2014.
The prostate-specific antigen (PSA) is a typical example of a biomarker with high diagnostic efficiency. In a meta-analysis published by Djulbegovic et al. in 2010 and including 6 randomized controlled trials with a total of 387,286 participants, PSA screening was associated with a significantly increased probability of being diagnosed with prostate cancer (relative risk [RR], 1.46; 95% confidence interval [CI], 1.21–1.77) and stage I prostate cancer (RR, 1.95; 95% CI, 1.22–3.13), although no effect was found for mortality from prostate cancer (RR; 0.88; 95% CI, 0.71–1.09) and overall deaths (RR, 0.99; 95% CI, 0.97–1.01). These results were confirmed in a more recent meta-analysis published by Ilic et al. in 2013, and including 5 randomized controlled trials with a total of 341,342 participants. Although the diagnosis of prostate cancer was significantly more common in men randomized to PSA screening (RR, 1.30; 95% CI, 1.02–1.65), no significant difference was found in prostate cancer mortality between men randomized to PSA screening and the control group (RR, 1.00; 95% CI, 0.86–1.17). Although a further improvement in diagnostic efficiency can be obtained by lowering the cut-off or adopting surrogate approaches (ie, age-specific reference ranges, PSA density, PSA velocity, percent-free PSA, or Prostate Health Index),

It is hence evaluated by means of receiver operating characteristic curves, sensitivity, specificity, predictive value (negative or positive), likelihood ratio (negative or positive), number needed to diagnose (ie, the number of tests that need to be performed to gain a positive response for the presence of disease), diagnostic odds ratio (ie, the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease) or reclassification (ie, the proportion of individuals correctly reclassified between risk categories). The diagnostic efficacy can instead be assessed by performing randomized trial and survival analysis (eg, the Cox model), which investigate the time between biomarker implementation and a subsequent event or endpoint (ie, complications or death). As such, it is rather understandable that the terms efficiency and efficacy cannot (and should not) be used interchangeably, although they frequently are in published research.

**Diagnostic efficiency, clinical efficacy, or both?** According to the previous definitions, 3 different scenarios emerge, in which a given biomarker can display a high degree of diagnostic efficiency, a high degree of clinical efficacy, or a high degree of both (FIGURE 3).
In particular, Reichlin et al. performed a multicenter study to investigate the diagnostic efficiency of 4 HS troponin immunoassays compared with a standard method, and found that the area under the curve of the novel immunoassays was significantly higher than that of the standard method (ie, from 0.95–0.96 vs 0.90). These results were then confirmed in a meta-analysis by Li et al., who concluded that HS troponin immunoassays exhibit a much higher sensitivity than conventional methods (0.79 vs 0.59), with comparable specificity (0.92 vs 0.95). Even more importantly, this greater diagnostic efficiency was also translated into larger clinical efficacy, since an additional meta-analysis showed that elevated troponin values measured with HS immunoassays conferred a significantly higher risk of all-cause mortality compared with normal values (odds ratio [OR], 4.35; 95% CI; 2.81–6.73).

In another systematic review and meta-analysis of the literature including 27 studies in noncardiac patients, Ahmed et al. also reported that an increased troponin value was associated with increased in-hospital and 30-day mortality (OR, 3.88; 95% CI, 2.90–5.19), as well as with an increased risk of long-term mortality at 6 months (OR, 4.21; 95% CI, 1.84–9.64). According to these data, and at variance with PSA and RDW, cardiac-specific troponins should hence be regarded as ideal biomarkers that fulfill the criteria of both diagnostic efficiency and clinical efficacy (Figure 3).

The red blood cell distribution width (RDW) is a simple and inexpensive measure of anisocytosis (ie, heterogeneity of erythrocyte volume). Growing evidence indicates that this parameter is strongly associated with the prognosis (ie, complications and overall death) of a large number of human disorders. Interestingly, increased RDW values have been convincingly associated with unfavorable outcome in patients with cardiovascular disease, venous thrombosis, diabetes, cancer, kidney and liver disease, as well as in critically ill patients and in the general population. It is hence noteworthy that the RDW has a meaningless diagnostic efficiency for a given disease (ie, poor specificity), since its value is increased in all those conditions that alter erythropoiesis or impair erythrocyte survival, but has a high clinical efficacy since its measurement would allow to identify a subset of patients who may benefit from a more intensive management in order to improve clinical outcomes.

A third paradigmatic example of a biomarker displaying a high degree of both diagnostic efficiency and clinical efficacy is that of cardiac-specific troponins. A novel generation of troponin immunoassays, conventionally known as HS, has recently been developed and commercialized. These methods are typically characterized by an improved analytical sensitivity, which has allowed to consistently decrease the diagnostic thresholds for acute coronary syndrome, up to undetectable levels of these biomarkers.

The overall improvement of diagnostic efficiency for diagnosing acute coronary syndrome has been confirmed in a large number of studies. In particular, Reichlin et al. performed a multi-center study to investigate the diagnostic efficiency of 4 HS troponin immunoassays compared with a standard method, and found that the area under the curve of the novel immunoassays was significantly higher than that of the standard method (ie, from 0.95–0.96 vs 0.90). These results were then confirmed in a meta-analysis by Li et al., who concluded that HS troponin immunoassays exhibit a much higher sensitivity than conventional methods (0.79 vs 0.59), with comparable specificity (0.92 vs 0.95). Even more importantly, this greater diagnostic efficiency was also translated into larger clinical efficacy, since an additional meta-analysis showed that elevated troponin values measured with HS immunoassays conferred a significantly higher risk of all-cause mortality compared with normal values (odds ratio [OR], 4.35; 95% CI; 2.81–6.73).

In another systematic review and meta-analysis of the literature including 27 studies in noncardiac patients, Ahmed et al. also reported that an increased troponin value was associated with increased in-hospital and 30-day mortality (OR, 3.88; 95% CI, 2.90–5.19), as well as with an increased risk of long-term mortality at 6 months (OR, 4.21; 95% CI, 1.84–9.64).

According to these data, and at variance with PSA and RDW, cardiac-specific troponins should hence be regarded as ideal biomarkers that fulfill the criteria of both diagnostic efficiency and clinical efficacy (Figure 3).

Drawbacks of translating diagnostic efficacy into clinical efficacy Although the promise of biomarkers is clear, significant drawbacks emerge in the challenging process of translating diagnostic efficacy into clinical efficacy (Table). First, a rather feeble relationship exists between biomarker research and the leading causes of death worldwide. Interestingly, less than 2% of biomarker studies in the past few years involved highly prevalent diseases such as lower respiratory infections, chronic
obstructive pulmonary disease, or diarrhea, which cumulatively average a much greater mortality rate than ischemic heart disease or cerebrovascular disorders, 2 pathologies with 20-fold more articles published.19 As such, the major focus should be placed on bridging the gap between basic research and clinical practice to develop high-quality clinically-driven research on the most important strategic challenges facing health care systems.20

The second important issue is the effective translation of biomarker research from the bench to the bedside. Despite considerable efforts, there is now unquestionable evidence that very few putative biomarkers have been validated and successfully integrated into daily clinical practice.21 The number of biomarkers that were “lost in translation” is remarkably high, much greater than 70%. This is due to a number of reasons including inadequateness of derivation and validation cohorts, low accuracy, poor analytical performance (ie, imprecision, repeatability, reproducibility, linearity), analytical sensitivity (ie, limit of blank, limit of detection, and limit of quantification) that is too scarce to be clinically useful, poor specificity, matrix effect, analyte instability, insufficient transparency and disclosure in published material,22 as well as errors or fraud in published data that led to retraction of a substantial number of papers.23

The biological variability is another important issue. The within-subject biological variability is defined as variation within a single individual estimated as a pooled variation from a group of individuals, whereas the between-subject biological variability is defined as variation between the central tendencies of a group of individuals.24 Rather understandably, the clinical efficacy of a novel biomarker is strongly influenced by the biological variability. Biomarkers with very high within-subject biological variation cannot be efficiently used for patient monitoring, since their physiological variation may be close (or even exceed) that caused by disease progression or improvement. On the other hand, the problems with biomarkers displaying very high between-subject biological variation mainly emerge from the challenging identification of reliable reference range and diagnostic cut-off that could be widely used for screening or diagnosis. In such case, the most reliable approach entails population partitioning into many separate and homogenous classes (eg, age, sex, race, body mass index, etc.), but this is inherently a long, challenging, and expensive enterprise.

Technology drawbacks may represent an important limitation in biomarker development pipeline. Most circulating biomarkers are discovered using sophisticated proteomic technologies, which are unsuitable for most clinical laboratories. The further development of commercial assays is not a trivial endeavor and requires a considerable amount of time and huge economic resources. Accordingly, the expected outcome, that is, the final development of a commercial method, does not come for free, since the final price of the test will be strongly influenced by the efforts put in research and development, as well as by the specific technique (eg, the price of an enzyme-linked immunoassay is many orders of magnitude greater than that of a simple colorimetric or enzymatic method). The inadequateness of reimbursement rates may represent an additional problem, especially when the cost of the test is close to or even greater than the level of reimbursement.

It is intuitively appealing that any new diagnostic test displaying excellent diagnostic efficacy and high clinical efficacy would make complete the validation phase, to be routinely implemented in the vast majority of clinical laboratories. Despite the technological issues outlined in the previous paragraph, the introduction of novel biomarkers into clinical practice inevitably requires an accurate cost-effective analysis.25 In a world with limited resources, still plagued by an unprecedented economic crisis, magnification of laboratory panels is not always sustainable and affordable. It should hence be necessary to prove that the introduction of a novel biomarker is effective in improving clinical outcomes sufficiently to justify the additional costs of testing and treatment.

According to recent statistics, approximately 30% of United States health care expenditure is considered as duplicative or unnecessary. Even more importantly, as many as 300 billion USD are wasted each year in the United States for inappropriate use or overutilization of medical tests.26 This is noteworthy since the inappropriate usage of laboratory resources not only compromises the quality of care, but may also generate tangible risks or harm to patients by leading to more testing and unnecessary procedures or medication.27 This evidence is challenging the minds of a large number of laboratory professionals and healthcare administrators, since the inherent risk of inappropriateness and further unjustified healthcare expenditure would ultimately pose serious psychological hurdles to the implementation of innovative and potentially effective tests.

The final issue is represented by the role of regularity agencies in the biomarker development pipeline. Although the regulatory frameworks for regulation of in-vitro diagnostic devices (ie, the Federal Food and Drugs Administration in the United States and directive 98/79/EC in the European Community) may differ across different countries, there is convincing evidence that the coordinated activities of these regularity agencies may be indeed effective to drive quality improvements. However, the risk that over-regulation will affect test availability is tangible, if not virtually unavoidable, for the so called “in-house” testing,28 namely, in-vitro diagnostic tests that use research kits, combinations of reagents developed in-house, or commercially developed tests used for indications other than those for which they are officially marked.
Conclusions The discovery of biomarkers is expanding at an unprecedented rate and no decrease in publications is expected. The vast majority of published studies on biomarkers have been mainly focused on diagnostic efficacy rather than on clinical efficacy, and this seems no longer sustainable and affordable in a world with limited resources. The laboratory community should hence place more focus on what the clinicians really want and need, rather than wasting valuable human and economic resources for pursuing an endless search for the analytically perfect test. This would require a paradigm shift in biomarker research, wherein the development and commercialization of novel biomarkers should be guided by efficacy studies and not be limited to a simple analysis of diagnostic efficiency. This implies discriminating (with good accuracy) between patients with favorable prognosis and those with unfavorable outcomes, and then describing how a novel test can effectively modify the natural history of disease.

In brief, the key points to bridge the gap between efficiency and efficacy include a preliminary “proof of concept”, wherein the exploratory biomarker is evaluated for outcomes and not against its diagnostic performance. This would obviously require planning and carrying out randomized trials of diagnostic tests or screening strategies as commonplace for trials of drugs and biomedical devices. Another important aspect is the link between biomarker and pathophysiology. In other words, the new biomarker should target one or more specific pathways that play an essential role in the natural history of disease and that can be successfully modified with preventive or therapeutic interventions. Expectedly, the clinical information emerging from the introduction of a novel test should also be completely independent, or at least more accurate, than that provided by existing biomarkers. This concept is basically defined as “incremental validity” (i.e., the degree to which a measure explains or predicts some phenomena of interest, relative to other measures). It has been in fact demonstrated that improving the performance of diagnostic models may be possible using a discrete number of biomarkers, provided that these are weakly correlated or uncorrelated with each other. In the foreseeable future, efficacy should hence be improved by translation of these concepts into a simple “six R” paradigm, namely, performing the Right test, with the Right method, at the Right time, to the Right patient, at the Right cost, and for obtaining the Right clinical outcome.

REFERENCES

26 Hilborne L. Choosing wisely: selecting the right test for the right patient at the right time. MLO Med Lab Obs. 2014; 46: 40.
ARTYKUŁ POGŁĄDOWY

Paradygmat biomarkerów: między efektywnością diagnostyczną a skutecznością kliniczną

Giuseppe Lippi¹, Camilla Mattiuzzi²

¹ Laboratory of Clinical Chemistry and Hematology, Academic Hospital of Parma, Parma, Włochy
² Service of Clinical Governance, General Hospital of Trento, Trento, Włochy

SŁOWA KLUCZOWE
biomarkery,
dokładność,
efektywność,
skuteczność

STRESZCZEŃE

Zainteresowanie badaniami biomarkerów wzrasta lawinowo i nie przewiduje się, aby ten trend wkrótce uległ odwróceniu. Mimo że przydatność kliniczna testów laboratoryjnych jest zwykle definiowana w kategorii efektywności diagnostycznej lub skuteczności klinicznej, pojęcia te uzupełniają się wzajemnie, ale nie są wymienne. Pierwsze z tych pojęć stanowi wyraz dokładności diagnostycznej, ale nie pociąga za sobą oceny wyniku. Skuteczność kliniczna ocenia natomiast, czy pewne badania mogą spowodować znaczące zmiany w opiece i poprawę wyników klinicznych. Zdecydowana większość opublikowanych badań koncentrowała się głównie na skuteczności diagnostycznej, a nie na skuteczności klinicznej – jak się wydaje, w świecie o ograniczonych zasobach nie jest to zrównoważone podejście. Wypełnienie luki pomiędzy efektywnością a skutecznością nie jest blahym przedsięwzięciem, dlatego konieczna będzie zmiana paradygmatu: pracownicy laboratoriów powinni skupiać się na tym, czego potrzebują lekarze, a nie prowadzić niekończące się poszukiwania analitycznie doskonałych testów. W dającej się przewidzieć przyszłości należy poprawić skuteczność, sprowadzając to pojęcia do paradygmatu "sześciu W", tj. wykonanie Właściwego testu, Właściwą metodą, we Właściwym czasie, Właściwemu pacjentowi, za Właściwą cenę oraz uzyskanie Właściwego wyniku.